

# Clustering of non-ergodic eigenstates in quantum spin glasses

C. L. Baldwin,<sup>1,2</sup> C. R. Laumann,<sup>1</sup> A. Pal,<sup>3</sup> and A. Scardicchio<sup>4,5</sup>

<sup>1</sup>*Department of Physics, Boston University, Boston, MA 02215, USA*

<sup>2</sup>*Department of Physics, University of Washington, Seattle, WA 98195, USA*

<sup>3</sup>*Rudolf Peierls Centre for Theoretical Physics, Oxford University, Oxford OX1 3NP, UK*

<sup>4</sup>*Abdus Salam ICTP Trieste, Strada Costiera 11, 34151 Trieste, Italy*

<sup>5</sup>*INFN, Sezione di Trieste, Via Valerio 2, 34127 Trieste, Italy*

(Dated: November 9, 2016)

The two primary categories for eigenstate phases of matter at finite temperature are many-body localization (MBL) and the eigenstate thermalization hypothesis (ETH). We show that in the paradigmatic quantum  $p$ -spin models of spin-glass theory, eigenstates violate ETH yet are not MBL either. A mobility edge, which we locate to leading order in  $1/p$  using the forward-scattering approximation and replica techniques, separates the non-ergodic phase at small transverse field from an ergodic phase at large transverse field. The non-ergodic phase is also bounded from above in temperature, by a transition in configuration-space statistics reminiscent of the clustering transition in spin-glass theory. We show that the non-ergodic eigenstates are organized in clusters which exhibit distinct magnetization patterns, as characterized by an eigenstate variant of the Edwards-Anderson order parameter.

Many systems under experimental investigation as platforms for many-body localization (MBL) [1–9] have long-range interactions that mediate the direct transport of excitations. This includes disordered electronic materials [10, 11], ion traps [12], interacting NV centers in diamond [13, 14], and superconducting qubit devices developed for adiabatic quantum computing [15–17]. In sufficiently long-ranged systems, the proliferation of long-distance resonances precludes quantum mechanical localization [18–22], an intuitive result strongly supported by analytic work over the last half century. Nevertheless, the quantum Random Energy Model (QREM), an infinite-range spin glass, was recently shown to exhibit a phase with localized eigenstates at finite energy density [23, 24]. The QREM provides an analytically tractable framework for studying mobility edges and configuration-space localization. This raises the obvious question of how localization survives despite the infinite-range interactions and what role it plays in more realistic long-range systems.

Some insight comes from considering the distribution of local fields — i.e., the energy required to flip one of the system’s  $N$  spins relative to a given configuration. In the QREM, flipping a spin typically changes the energy by  $O(N)$ . Thus the quantum fluctuations which lead to the proliferation of resonances are strongly suppressed. However, short-range models have  $O(1)$  local fields, and in fact, so do power-law and infinite-range systems with general  $p$ -body interactions. This suggests that the eigenstate-localized phase of the QREM is an exceptional case among long-range models: strict configuration-space localization cannot exist in any model with  $O(1)$  local fields, since the introduction of quantum dynamics causes resonant fluctuations.

In this paper, we study the eigenstate properties of the quantum  $p$ -spin models [25–28], a family of infinite-range quantum spin glasses. These models have  $O(1)$  local fields and reproduce the QREM in the  $p \rightarrow \infty$  limit.

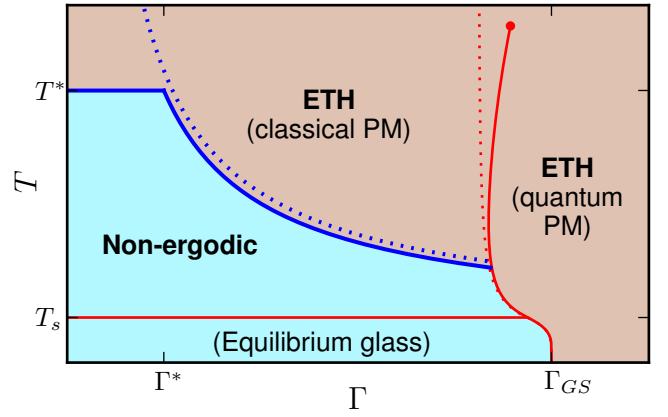


FIG. 1. The  $T - \Gamma$  phase diagram of the quantum  $p$ -spin model, sketched for arbitrary  $p$ . The indicated  $T$  and  $\Gamma$  values scale as:  $T_s = O(1)$ ,  $T^* \simeq \sqrt{\frac{p}{4 \ln p}}$ ,  $\Gamma^* \simeq \sqrt{\frac{\ln p}{p}}$ ,  $\Gamma_{GS} = O(1)$ . Solid red lines are phase boundaries obtained through the imaginary-time replica formalism [25], and the solid blue line is the eigenstate phase boundary. The corresponding dashed red and blue lines are those of the QREM ( $p \rightarrow \infty$  limit).

At large but finite  $p$ , we find that the QREM’s phase of strictly localized eigenstates gives way to a phase of partially delocalized yet non-ergodic eigenstates (blue in Fig. 1), similar to what was observed in the context of single-particle localization on the Bethe lattice [29–31]. This is in contrast to the fully-delocalized paramagnetic phase (orange in Fig. 1), in which eigenstates satisfy the eigenstate thermalization hypothesis (ETH) [32–34] and thus exhibit thermal behavior.

The formal distinction that we make between ergodic, non-ergodic, and MBL eigenstates concerns the off-diagonal matrix elements of local operators between them. Schematically, denote by  $\hat{\sigma}$  any operator supp-

ported on  $O(1)$  spins and consider the state  $\hat{\sigma}|\Psi\rangle$  (with  $|\Psi\rangle$  an eigenstate). According to ETH [35], the overlap with any other eigenstate  $|\Phi\rangle$  at the same energy density  $\epsilon$  should scale as  $\langle\Phi|\hat{\sigma}|\Psi\rangle \sim \frac{1}{\sqrt{\exp(Ns_{\text{eq}}(\epsilon))}} f(\epsilon, E_\Psi - E_\Phi)$ , where  $s_{\text{eq}}(\epsilon)$  is the thermodynamic entropy density and  $f$  is a smooth function of  $\epsilon$  and the energy difference. Our analysis below suggests that the eigenstates of the ETH phase in Fig. 1 obey this scaling. On the other hand, in an MBL phase  $\hat{\sigma}|\Psi\rangle$  should have significant weight only on  $O(1)$ -many eigenstates [2], a notion one can make precise through a participation ratio (e.g.,  $\sum_\Phi |\langle\Phi|\hat{\sigma}|\Psi\rangle|^4 \sim O(1)$ ). We find that the eigenstates of the non-ergodic phase *do not* obey this definition of MBL, even though they violate ETH. Rather, they are organized into “clusters” (defined below). Within a cluster  $c$ , eigenstates follow ETH-type scaling:

$$\langle\Phi^{(c)}|\hat{\sigma}|\Psi^{(c)}\rangle \sim \frac{1}{\sqrt{\exp(Ns_c(\epsilon))}} f_c(\epsilon, E_\Psi - E_\Phi), \quad (1)$$

but off-diagonal matrix elements between clusters are heavily suppressed ( $c \neq c'$ ),

$$\langle\Phi^{(c')}|\hat{\sigma}|\Psi^{(c)}\rangle \ll \frac{1}{\sqrt{\exp(Ns_{\text{eq}}(\epsilon))}}. \quad (2)$$

Here,  $|\cdot^{(c)}\rangle, |\cdot^{(c')}\rangle$  are eigenstates belonging to each cluster.  $s_c(\epsilon)$  is the entropy density within  $c$  (which is strictly less than  $s_{\text{eq}}(\epsilon)$ ), and  $f_c$  is a smooth, cluster-dependent  $O(1)$  function. More physically, such non-ergodic eigenstates are thermal within a cluster but not thermal in configuration space as a whole.

Concretely, the quantum  $p$ -spin models are defined by

$$H_p = - \sum_{(i_1 \dots i_p)} J_{i_1 \dots i_p} \hat{\sigma}_{i_1}^z \dots \hat{\sigma}_{i_p}^z - \Gamma \sum_{i=1}^N \hat{\sigma}_i^x \equiv H_p^C + H^Q, \quad (3)$$

where the classical term  $H_p^C$  sums over all distinct  $p$ -tuples of  $N$  spins and the quantum term  $H^Q$  provides a uniform transverse field. From the configuration-space perspective,  $H_p^C$  is the random potential and  $H^Q$  hops on the edges of the hypercube. The random couplings  $J_{i_1 \dots i_p}$  are i.i.d. Gaussians of mean 0 and variance  $\frac{p!}{2N^{p-1}}$ , to ensure extensivity. These models, and in particular the Sherrington-Kirkpatrick model ( $p = 2$ ) [36, 37], feature prominently in the theory of mean-field spin glasses [38–41]. It is known [25] that the thermodynamic free energy of  $H_p$  approaches that of the QREM as  $p$  increases. As we will see, the eigenstate phases of the  $p$ -spin model also approach those of the QREM in this limit. In order to show this, we study the eigenstates within the forward-scattering approximation (FSA) [42], and determine the leading-in- $\frac{1}{p}$  behavior.

Before we turn to the detailed analysis, it is useful to consider the  $p$ -spin models in terms of Anderson localization on the  $N$ -dimensional hypercube defined by the  $\sigma^z$

configuration space. The QREM corresponds to a hopping model in an uncorrelated Gaussian random potential of bandwidth  $\sqrt{N}$ . This bandwidth models that of a many-body system with extensive spectrum, but the lack of correlations implies unrealistically large local fields. In the  $p$ -spin model, the potential remains Gaussian but exhibits correlations which restrict the energy differences between adjacent sites to be  $O(p)$ . This leads to the entropically large clusters over which the eigenstates delocalize at short fractional Hamming distance (see below). The phase transition at finite transverse field shown in Fig. 1 corresponds to eigenstates tunneling *between* clusters.

To obtain the perturbative eigenstates we must first understand the correlations in the classical energy landscape in more detail. At Hamming distance  $Nx$  from a given configuration with energy  $N\epsilon_0$ , the random potential obeys a conditional distribution [38],

$$P_x(\epsilon) \propto \exp\left(-N \frac{(\epsilon - (1-2x)^p \epsilon_0)^2}{1 - (1-2x)^{2p}}\right). \quad (4)$$

At distance  $x = \frac{1}{N}$ , corresponding to a single spin flip, this expression provides the distribution of local fields. It is easy to check that these are indeed  $O(p)$  and that  $P_{\frac{1}{N}}(\epsilon_0) > 0$  even in the thermodynamic limit.

Using Eq. (4), the average number of states at fractional distance  $x$  with energy density  $\epsilon$  matching  $\epsilon_0$  is  $\binom{N}{Nx} P_x(\epsilon_0) \sim e^{Ns(x)}$ , with

$$s(x) = -x \ln x - (1-x) \ln(1-x) - \frac{1 - (1-2x)^p}{1 + (1-2x)^p} \epsilon_0^2. \quad (5)$$

In order to compare with the literature, it is useful to parametrize  $\epsilon_0$  through the temperature  $T$  defined by formal Legendre transform, even when the system fails to thermalize dynamically. It is shown in the Appendix that  $\epsilon_0 = -\frac{1}{2T} + O\left(\frac{1}{p^2}\right)$ . Eq. (30) is an annealed average which provides a rigorous upper bound for the typical number of states, since  $\mathbb{E}[\ln \dots] \leq \ln \mathbb{E}[\dots]$ . When  $s(x) < 0$ , we know with certainty that there are no configurations at  $x$  with energy density  $\epsilon_0$ .

The entropy  $s(x)$  is plotted in Fig. 2 for  $p = 6$  as illustration (we show only  $x \in [0, \frac{1}{2}]$ , since  $s(1-x) = s(x)$ ). A transition occurs at the temperature  $T^*$ . At  $T > T^*$ , there are configurations over the entire range of  $x$ , whereas at  $T < T^*$ , there is a “forbidden” region  $(x^*(T), x^{**}(T))$  in which no configurations lie. We have rigorously confirmed the presence of three distinct regions via a second-moment analysis, analogous to that done in Ref. [43]. See the Appendix for details. Thus the configurations at energy  $\epsilon_0$  form disconnected clusters of Hamming size  $x^*(T)$ . Energy levels within a cluster are highly correlated, but those of different clusters are essentially independent (see Eq. (4)). This behavior is analogous to the clustering phenomenon observed in a variety of other settings, e.g.,  $k$ -SAT problems [43], coloring of random graphs [44], and in particular, the replica theory of spin glasses (more discussion below).

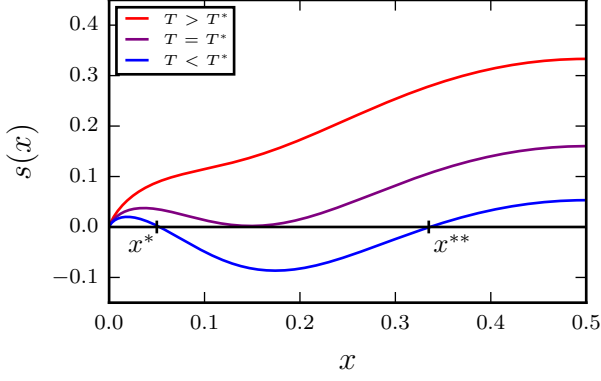


FIG. 2. The annealed Hamming-distance-resolved entropy  $s(x)$ . Conditioning on the configuration at  $x = 0$  having energy density  $\epsilon_0 = -\frac{1}{2T}$ ,  $e^{Ns(x)}$  is the average number of states at Hamming distance  $Nx$  with energy density  $\epsilon_0$ . These curves are for the *classical*  $p$ -spin model (i.e.,  $\Gamma = 0$ ) with  $p = 6$ . The specific temperatures used are, from reddest curve to bluest curve,  $\{T = 0.83, T = 0.69, T = 0.63\}$ .

By setting  $s(x) = \partial_x s(x) = 0$ , we find that

$$T^* = \sqrt{\frac{p}{4 \ln p}} \left( 1 + O\left(\frac{\ln \ln p}{\ln p}\right) \right). \quad (6)$$

Note that  $T^* \rightarrow \infty$  as  $p \rightarrow \infty$ . Furthermore, at  $T \ll T^*$ ,  $x^*(T) \sim e^{1 - \frac{T^*2}{T^2} \ln p}$  whereas  $x^{**}(T) \sim O(1)$  (with respect

to  $p$ ). Thus the clusters below  $T^*$  are well separated, and they become arbitrarily small as  $p$  increases. Regardless, the clusters cover a macroscopic Hamming distance at any finite  $p$ .

With this understanding of how the *classical* states at energy density  $\epsilon_0$  are organized, we now introduce the transverse field  $\Gamma$  and study eigenstates using the FSA [42]. The eigenstate  $|\Psi_\alpha\rangle$  may be expressed as a perturbation series in  $\Gamma$  starting from a classical state  $|\alpha\rangle$ . The amplitude on a different classical state  $|\beta\rangle$  is a sum over the directed sequences (“paths”) of spin flips that transform  $\alpha$  into  $\beta$ :

$$\langle \beta | \Psi_\alpha \rangle = \frac{\Gamma}{E_\alpha - E_\beta} \sum_{\mathcal{P}} \prod_{j=1}^{Nx_{\alpha\beta}-1} \frac{\Gamma}{E_\alpha - E_{\mathcal{P}_j}}, \quad (7)$$

where  $x_{\alpha\beta}$  is the fractional Hamming distance between  $\alpha$  and  $\beta$ , the sum is over the  $(Nx_{\alpha\beta})!$  sequences, and  $\mathcal{P}_j$  is the  $j$ ’th configuration along sequence  $\mathcal{P}$ . See Ref. [24] for a more detailed explanation.

If  $|\langle \beta | \Psi_\alpha \rangle| \sim 1$ , we say that  $\beta$  and  $\alpha$  are resonant. Knowing where resonances occur in configuration space tells us where the eigenstate amplitude is concentrated. We show in the Appendix that the *typical* path amplitude to a configuration at distance  $Nx$  is of the form  $e^{-Nk(x)}$ . A replica calculation (also in the Appendix), which treats  $|\langle \beta | \Psi_\alpha \rangle|$  as the partition function for a directed random polymer, suggests that the sum in Eq. (7) is dominated collectively by typical paths and not by rare fluctuations. Thus the expected number of resonances at  $x$  is  $e^{Nf(x)}$  with

$$f(x) = s(x) + \frac{1}{N} \ln(Nx)! - k(x) = x \ln \frac{\Gamma}{e|\epsilon_0|} - (1-x) \ln(1-x) - \int_0^x dy \ln(1 - (1-2y)^p) - \frac{1 - (1-2x)^p}{1 + (1-2x)^p} \epsilon_0^2. \quad (8)$$

Fig. 3 shows the three different qualitative behaviors of  $f(x)$  as  $T$  and  $\Gamma$  are varied. First consider  $T \ll T^*$ . There are resonances at small enough  $x$  (see the bottom panel of Fig. 3) which belong to the same cluster as the unperturbed state. These are inevitable. Regardless of how small  $\Gamma$  is, a non-zero fraction of spins will have much smaller local fields and resonate. On the other hand, resonances belonging to different clusters only appear when  $\Gamma$  exceeds a critical  $\Gamma_c(T)$ .

In principle, we should treat the intra-cluster states within degenerate perturbation theory and then perturb the hybridized states. We expect that this will cause further resonances in the initial cluster, until eventually every configuration in the cluster has hybridized. However, since they cannot extend past  $x^* \ll O(1)$ , this effect only gives subleading corrections to the number of resonances in other clusters. For  $\Gamma < \Gamma_c(T)$ , the eigenstate amplitude outside of its cluster is exponentially suppressed relative to inside. This leads us to assert that an ex-

citation, e.g.,  $\hat{\sigma}|\Psi\rangle$ , has weight distributed chaotically amongst those eigenstates within  $|\Psi\rangle$ ’s cluster but has much less weight elsewhere. The precise statement of this idea is Eq. (1).

Conversely, the inter-cluster hybridization that occurs for  $\Gamma > \Gamma_c(T)$  does significantly affect the wavefunction. Since different clusters have essentially independent energies, we expect that such resonances lead to more in other clusters, until the wavefunction ultimately extends throughout the entire configuration space. Off-diagonal matrix elements are then expected to obey ETH. We also cannot neglect the hybridizations within a cluster if the separation between clusters is comparable to their lengths, i.e., if  $T \gtrsim T^*$ , and we expect these eigenstates to be ergodic as well. Thus the red and green curves in Fig. 3 do not accurately represent their respective eigenstates. It remains an open problem to calculate these states and check the predictions of ETH.

Regardless, we can determine  $\Gamma_c(T)$  to leading or-

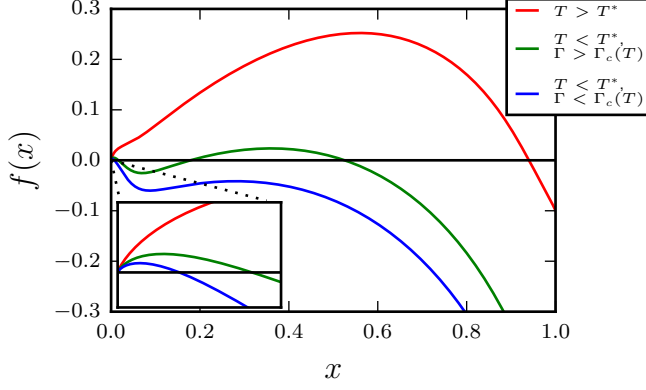


FIG. 3. The annealed entropy of resonances at distance  $x$ ,  $f(x)$ .  $e^{Nf(x)}$  is the average number of states at Hamming distance  $Nx$  which are resonant with an unperturbed state at  $x = 0$  having energy density  $\epsilon_0 = -\frac{1}{2T}$ . The inset shows  $f(x)$  at very small  $x$ . These curves are for  $p = 20$ . The red curve has  $(T, \Gamma) = (2.27, 0.50)$ , the green curve has  $(T, \Gamma) = (1.56, 0.50)$ , and the blue curve has  $(T, \Gamma) = (1.39, 0.50)$ .

der. By requiring that the maximum of  $f(x)$  over all  $x > x^*(T)$  be 0 (see Appendix), we obtain the portion of the blue curve below  $T^*$  in Fig. 1. To within  $O(\frac{1}{p})$  corrections, this is identical to that of the QREM [24]. Since  $T^*$  diverges as  $p$  increases, we find that the non-ergodic phase of  $H_p$  does indeed map continuously onto the MBL phase of  $H_{\text{QREM}}$ .

Furthermore, since the short-distance resonances occupy a small corner of configuration space, they do not affect observable properties commonly associated with MBL. One prominent observable in spin-glass theory is the Edwards-Anderson order parameter  $q_{\text{EA}} \equiv \frac{1}{N} \sum_i \langle \sigma_i \rangle^2$ , where the average is with respect to the Gibbs distribution. We define an eigenstate variant  $q_{\text{ES}}(\Psi) \equiv \frac{1}{N} \sum_i \langle \Psi | \sigma_i^z | \Psi \rangle^2$ . Note that  $q_{\text{ES}}(\Psi) = q_{\text{EA}}$  whenever ETH holds. Heuristically,  $q_{\text{ES}}(\Psi)$  measures how similar the configurations are for which  $|\Psi\rangle$  has significant amplitude.  $q_{\text{ES}}(\Psi) \sim 1$  means that measuring the  $\sigma^z$  configuration within state  $|\Psi\rangle$  will consistently give macroscopically similar results. One can then associate a specific magnetization pattern to  $|\Psi\rangle$ . As shown in the Appendix, in the non-ergodic phase of  $H_p$ ,

$$q_{\text{ES}} = 1 - \frac{4\Gamma^2 T^2}{p^2} + \dots \quad (9)$$

Compare to ergodic eigenstates in the paramagnetic phase, which have  $q_{\text{ES}} = 0$ .

The level statistics in the non-ergodic phase is Poisson as well, just as in many-body-localized systems. Regardless of how the eigenstates hybridize within a cluster, they cannot do so over more than the total number of configurations in the cluster, which is  $e^{N \exp(-p\epsilon_0^2)}$  (see Appendix). These levels strongly repel and have GOE

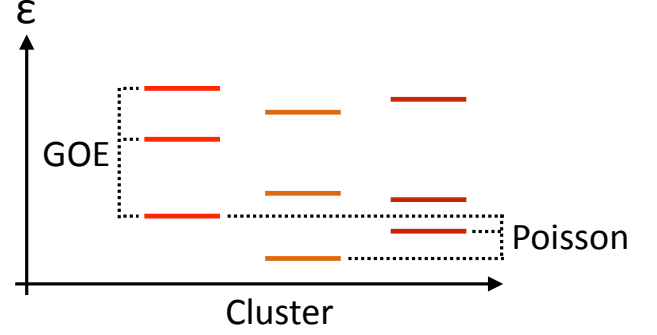


FIG. 4. A sketch of  $H_p$ 's eigenvalues in the non-ergodic phase (y-axis), organized by the cluster to which each eigenstate belongs (x-axis). Although the levels within a cluster are strongly correlated, the spectra of different clusters are independent and interpenetrating. Thus the spectrum as a whole has Poisson statistics.

statistics, but the spacing between them scales no smaller than  $e^{-N \exp(-p\epsilon_0^2)}$ . On the other hand, different clusters have independent fluctuations in energy levels, and the number of clusters is at least  $e^{N(\ln 2 - \epsilon_0^2 - \exp(-p\epsilon_0^2))} \gg e^{N \exp(-p\epsilon_0^2)}$ . The spectra of different clusters interpenetrate, so the level statistics is Poisson. This is sketched in Fig. 4.

Thus the FSA predicts that even though the many-body eigenstates of  $H_p$  are necessarily delocalized over an exponential number of configurations, they are highly non-ergodic at low  $T$  and small  $\Gamma$ . Numerical studies, e.g., exact diagonalization, could provide an important validation of this picture, yet one would need to study large  $p$  to separate the various features. The model loses its meaning if  $p$  is comparable to  $N$ , so one needs  $N \gg p \gg 1$  to conduct high-quality numerics. Current state-of-the-art techniques are limited to  $N \approx 30$  [45, 46], which is too small to separate these scales.

Regardless, the existence of clusters and a phase of non-ergodic eigenstates only requires the combinatorial arguments that we have described. Since this phase looks very similar to a many-body-localized phase in some regards, it raises the obvious question of how the two are related. The underlying physics is different: MBL is intrinsically a result of quantum interference, whereas the non-ergodic phase is more a consequence of  $O(N)$  energy and entropy barriers. In that respect, it relates more to the *classical* theory of glassiness in mean-field systems.

The relationship to mean-field spin-glass theory, and in particular the replica theory [47, 48], is potentially very deep. Most prominent is the connection between our  $T^*$  and the “dynamical” transition temperature  $T_d$  [27, 41, 49]. One interpretation of  $T_d$  is that below it, the Gibbs distribution concentrates around clusters in configuration space (although the equilibrium properties may still be paramagnetic). The heuristic picture is very similar to that of the transition we identify at  $T^*$ , and

although they are obtained through independent means, Eq. (6) for the asymptotic behavior of  $T^*$  agrees exactly with  $T_d$  in the literature [50]. The exact relationship between our calculations and the standard canonical analysis remains to be established. In Ref. [27] the authors studied  $H_p$  via replica theory and found an entire curve  $T_d(\Gamma)$ , which in other models was shown to relate to real-time dynamics in the presence of a heat bath [51]. That transition lies above the non-ergodic/ETH transition in Fig. 1; the connection between the two is an interesting open question.

On that note, it was recently argued [52] that the ergodicity of eigenstates need not imply ergodicity of dynamics. It is possible that the decay time of classical initial states might diverge in the thermodynamic limit when  $T < T^*$ , even when the corresponding eigenstates are ergodic. If this is so, the thermodynamic curve  $T_d(\Gamma)$  may describe the quench behavior of  $H_p$  rather than the eigenstate properties.

And finally, the intra-cluster structure of the non-ergodic eigenstates may be very rich. The ultrametric structure of Parisi's solution [53] suggests that a clus-

ter is organized into subclusters, which themselves have subclusters, and so on. The non-ergodic phase may actually be many different phases, with varying degrees of ergodicity-breaking corresponding to how many levels of clusters the eigenstates tunnel through, analogous to the physical picture of replica-symmetry-breaking.

*Acknowledgements* — We would like to thank L. Cugliandolo, G. Biroli, M. Sellitto, D. Huse, and A. Chandran for helpful discussions. We acknowledge the hospitality of ICTP, where part of this work was completed. C.L.B. also thanks the UW high-performance-computing club for providing necessary computer resources, and the NSF for support through a Graduate Research Fellowship, Grant No. DGE-1256082. C.R.L. acknowledges support from the Sloan Foundation through a Sloan Research Fellowship and the NSF through grant PHY-1520535. The work of A.P. was performed in part at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1066293. Note that any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

- 
- [1] B. Altshuler, Y. Gefen, A. Kamenev, and L. Levitov, *Physical Review Letters* **78**, 2803 (1997).
  - [2] D. Basko, I. Aleiner, and B. Altshuler, *Annals of physics* **321**, 1126 (2006).
  - [3] V. Oganesyan and D. A. Huse, *Phys. Rev. B* **75**, 155111 (2007).
  - [4] A. Pal and D. A. Huse, *Phys. Rev. B* **82**, 174411 (2010).
  - [5] M. Serbyn, Z. Papić, and D. A. Abanin, *Phys. Rev. Lett.* **111**, 127201 (2013).
  - [6] D. A. Huse, R. Nandkishore, and V. Oganesyan, *Phys. Rev. B* **90**, 174202 (2014).
  - [7] J. A. Kjäll, J. H. Bardarson, and F. Pollmann, *Phys. Rev. Lett.* **113**, 107204 (2014).
  - [8] M. Schreiber, S. S. Hodgman, P. Bordia, H. P. Lüschen, M. H. Fischer, R. Vosk, E. Altman, U. Schneider, and I. Bloch, *Science* **349**, 842 (2015), <http://science.sciencemag.org/content/349/6250/842.full.pdf>.
  - [9] A. Altland and T. Micklitz, *ArXiv e-prints* (2016), arXiv:1609.00877 [cond-mat.dis-nn].
  - [10] F. Ladieu, M. Sanquer, and J. P. Bouchaud, *Phys. Rev. B* **53**, 973 (1996).
  - [11] M. Ovadia, D. Kalok, I. Tamir, S. Mitra, B. Sacépé, and D. Shahar, *Scientific Reports* **5**, 13503 (2015).
  - [12] J. Smith, A. Lee, P. Richerme, B. Neyenhuis, P. W. Hess, P. Hauke, M. Heyl, D. A. Huse, and C. Monroe, *Nat Phys* **12**, 907 (2016).
  - [13] G. Kucsko, S. Choi, J. Choi, P. Maurer, H. Sumiya, S. Onoda, J. Isoya, F. Jelezko, E. Demler, N. Yao, and M. Lukin, *arXiv preprint arXiv:1609.08216* (2016).
  - [14] S. Choi, J. Choi, R. Landig, G. Kucsko, H. Zhou, J. Isoya, F. Jelezko, S. Onoda, H. Sumiya, V. Khemani, C. von Keyserlingk, N. Y. Yao, E. Demler, and M. D. Lukin, *ArXiv e-prints* (2016), arXiv:1610.08057 [quant-ph].
  - [15] M. W. Johnson, M. H. S. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk, E. M. Chapple, C. Enderud, J. P. Hilton, K. Karimi, E. Ladizinsky, N. Ladizinsky, T. Oh, I. Perminov, C. Rich, M. C. Thom, E. Tolkacheva, C. J. S. Truncik, S. Uchaikin, J. Wang, B. Wilson, and G. Rose, *Nature* **473**, 194 (2011).
  - [16] S. Boixo, T. F. Rønnow, S. V. Isakov, Z. Wang, D. Wecker, D. A. Lidar, J. M. Martinis, and M. Troyer, *Nature Physics* **10**, 218 (2014).
  - [17] S. Boixo, V. N. Smelyanskiy, A. Shabani, S. V. Isakov, M. Dykman, V. S. Denchev, M. H. Amin, A. Y. Smirnov, M. Mohseni, and H. Neven, *Nat Comm* **7**, 10327 (2016).
  - [18] P. W. Anderson, *Phys. Rev.* **109**, 1492 (1958).
  - [19] D. E. Logan and P. G. Wolynes, *The Journal of Chemical Physics* **87**, 7199 (1987).
  - [20] L. Levitov, *Annalen der Physik* **8**, 697 (1999).
  - [21] N. Y. Yao, C. R. Laumann, S. Gopalakrishnan, M. Knap, M. Müller, E. A. Demler, and M. D. Lukin, *Phys. Rev. Lett.* **113**, 243002 (2014).
  - [22] A. L. Burin, *Phys. Rev. B* **91**, 094202 (2015).
  - [23] C. R. Laumann, A. Pal, and A. Scardicchio, *Phys. Rev. Lett.* **113**, 200405 (2014).
  - [24] C. L. Baldwin, C. R. Laumann, A. Pal, and A. Scardicchio, *Phys. Rev. B* **93**, 024202 (2016).
  - [25] Y. Y. Goldschmidt, *Phys. Rev. B* **41**, 4858 (1990).
  - [26] T. M. Nieuwenhuizen and F. Ritort, *Physica A: Statistical Mechanics and its Applications* **250**, 8 (1998).
  - [27] L. F. Cugliandolo, D. R. Grempel, G. Lozano, and H. Lozza, *Phys. Rev. B* **70**, 024422 (2004).
  - [28] T. Jörg, F. Krzakala, J. Kurchan, and A. C. Maggs, *Phys. Rev. Lett.* **101**, 147204 (2008).
  - [29] G. Biroli, A. C. Ribeiro-Teixeira, and M. Tarzia, *ArXiv e-prints* (2012), arXiv:1211.7334 [cond-mat.dis-nn].
  - [30] A. De Luca, B. L. Altshuler, V. E. Kravtsov, and A. Scardicchio, *Phys. Rev. Lett.* **113**, 046806 (2014).
  - [31] B. L. Altshuler, E. Cuevas, L. B. Ioffe, and V. E.

- Kravtsov, Phys. Rev. Lett. **117**, 156601 (2016).
- [32] J. M. Deutsch, Phys. Rev. A **43**, 2046 (1991).
- [33] M. Srednicki, Phys. Rev. E **50**, 888 (1994).
- [34] M. Rigol, V. Dunjko, and M. Olshanii, Nature **481**, 224 (2008).
- [35] M. Srednicki, Journal of Physics A: Mathematical and General **32**, 1163 (1999).
- [36] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **35**, 1792 (1975).
- [37] G. Parisi, Phys. Rev. Lett. **43**, 1754 (1979).
- [38] B. Derrida, Phys. Rev. Lett. **45**, 79 (1980).
- [39] D. Gross and M. Mezard, Nuclear Physics B **240**, 431 (1984).
- [40] E. Gardner, Nuclear Physics B **257**, 747 (1985).
- [41] T. R. Kirkpatrick and D. Thirumalai, Phys. Rev. B **36**, 5388 (1987).
- [42] F. Pietracaprina, V. Ros, and A. Scardicchio, Phys. Rev. B **93**, 054201 (2016).
- [43] M. Mézard, T. Mora, and R. Zecchina, Physical Review Letters **94**, 197205 (2005).
- [44] R. Mulet, A. Pagnani, M. Weigt, and R. Zecchina, Phys. Rev. Lett. **89**, 268701 (2002).
- [45] D. J. Luitz, N. Laflorencie, and F. Alet, Phys. Rev. B **91**, 081103 (2015).
- [46] A. Leroze, V. K. Varma, F. Pietracaprina, J. Goold, and A. Scardicchio, arXiv preprint arXiv:1511.09144 (2015).
- [47] M. Mézard, G. Parisi, and M. Virasoro, Spin Glass Theory and Beyond, Lecture Notes in Physics Series (World Scientific Publishing Company, Incorporated, 1987).
- [48] M. Mézard and A. Montanari, Information, Physics, and Computation, Oxford Graduate Texts (OUP Oxford, 2009).
- [49] H. Sompolinsky and A. Zippelius, Phys. Rev. Lett. **47**, 359 (1981).
- [50] U. Ferrari, The dynamical transition of spin glasses with multi-body Ph.D. thesis, Universita di Roma (2012).
- [51] L. F. Cugliandolo, D. R. Grempel, and C. A. da Silva Santos, Phys. Rev. B **64**, 014403 (2001).
- [52] A. Chandran, A. Pal, C. Laumann, and A. Scardicchio, arXiv preprint arXiv:1605.00655 (2016).
- [53] M. Mézard, G. Parisi, N. Sourlas, G. Toulouse, and M. Virasoro, Phys. Rev. Lett. **52**, 1156 (1984).

## APPENDIX TO “CLUSTERING OF NON-ERGODIC EIGENSTATES IN QUANTUM SPIN GLASSES”

### A. Thermodynamics in the paramagnetic phase.

Before studying the eigenstates of the  $p$ -spin model, we describe the paramagnetic phase of the model’s thermodynamics. This analysis comes from Goldschmidt’s paper [25], in which he applies the (imaginary-time) path-integral replica formalism to this model. He also makes the static approximation, i.e., sets the imaginary-time spin autocorrelation to a constant. Yet since the paramagnetic phase is replica-symmetric, we do not include any discussion of replica-symmetry-breaking.

Starting from Eq. 19 in [25] with  $Q = \lambda = 0$ , the free energy is

$$f = -\frac{1}{4T}\chi^p + \frac{1}{2T}\chi^\nu - T \ln \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} 2 \cosh \frac{\sqrt{\Gamma^2 + \nu z^2}}{T}. \quad (10)$$

$\chi$  is the correlation  $\langle \sigma^z(k) \sigma^z(k') \rangle$  of a single spin between different imaginary times  $k$  and  $k'$ , and  $\nu$  is the associated Lagrange multiplier. It is standard to make the static approximation, in which  $\langle \sigma^z(k) \sigma^z(k') \rangle$  is assumed independent of  $k - k'$ . The equations that minimize  $f$  are

$$\nu = \frac{1}{2} p \chi^{p-1}, \quad (11)$$

$$\chi = T \left( \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} 2 \cosh \frac{\sqrt{\Gamma^2 + \nu z^2}}{T} \right)^{-1} \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \frac{z^2}{\sqrt{\Gamma^2 + \nu z^2}} 2 \sinh \frac{\sqrt{\Gamma^2 + \nu z^2}}{T}. \quad (12)$$

In the limit  $\nu \ll 1$ ,

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} 2 \cosh \frac{\sqrt{\Gamma^2 + \nu z^2}}{T} &= \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} 2 \cosh \frac{\Gamma}{T} \left( 1 + \frac{\nu z^2}{2\Gamma^2} + O\left(\frac{\nu^2}{\Gamma^4}\right) \right) \\ &= \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \left( 2 \cosh \frac{\Gamma}{T} \right) \left( 1 + \frac{\nu z^2}{2\Gamma T} \tanh \frac{\Gamma}{T} + O\left(\frac{\nu^2}{\Gamma^3 T} \left( 1 + \frac{\Gamma}{T} \right)\right) \right) \\ &= \left( 2 \cosh \frac{\Gamma}{T} \right) \left( 1 + \frac{\nu}{2\Gamma T} \tanh \frac{\Gamma}{T} + O\left(\frac{\nu^2}{\Gamma^3 T} \left( 1 + \frac{\Gamma}{T} \right)\right) \right). \end{aligned} \quad (13)$$

Concretely, this expansion is valid so long as  $\nu \ll \Gamma^2, \Gamma T$ .

In the limit  $\nu \gg 1$ ,

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} 2 \cosh \frac{\sqrt{\Gamma^2 + \nu z^2}}{T} &= 2 \int_0^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2 + \frac{\sqrt{\Gamma^2 + \nu z^2}}{T}} + O(1) \\
&= 2 \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(z + \frac{\sqrt{\nu}}{T}\right)^2 + \frac{\sqrt{\nu}}{T}\left(z + \frac{\sqrt{\nu}}{T}\right) \sqrt{1 + \frac{\Gamma^2}{\nu}\left(z + \frac{\sqrt{\nu}}{T}\right)^{-2}}} + O(1) \\
&= 2 \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{\frac{\nu}{2T^2} - \frac{1}{2}z^2 + \frac{\Gamma^2}{2T\sqrt{\nu}}\left(z + \frac{\sqrt{\nu}}{T}\right)^{-1} + O\left(\frac{\Gamma^4 T^2}{\nu^3}\right)} + O(1) \\
&= 2e^{\frac{\nu}{2T^2}} \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \left(1 + \frac{\Gamma^2}{2\nu} - \frac{\Gamma^2 T}{2\sqrt{\nu}^3} z + \frac{\Gamma^2 T^2}{2\nu^2} z^2 - \frac{\Gamma^2 T^3}{2\sqrt{\nu}^5} z^3 + O\left(\frac{\Gamma^4}{\nu^2}\right)\right) \\
&= 2e^{\frac{\nu}{2T^2}} \left(1 + \frac{\Gamma^2}{2\nu} + \frac{\Gamma^2 T^2}{2\nu^2} + O\left(\frac{\Gamma^4}{\nu^2}\right)\right).
\end{aligned} \tag{14}$$

In going from the first to the second line, we took  $z \rightarrow z + \frac{\sqrt{\nu}}{T}$ . This expansion is valid so long as  $\nu \gg T^2, \Gamma T$ . Thus

$$f \sim -\frac{1}{4T}\chi^p + \frac{1}{2T}\chi^\nu - \begin{cases} T \ln \left(2 \cosh \frac{\Gamma}{T}\right) + \frac{\nu}{2T} \tanh \frac{\Gamma}{T}, & \nu \ll \Gamma^2, \Gamma T \\ T \ln 2 + \frac{\nu}{2T} + \frac{\Gamma^2 T}{2\nu} + \frac{\Gamma^2 T^3}{2\nu^2}, & \nu \gg T^2, \Gamma T \end{cases} \tag{15}$$

The saddle-point equation (12) (i.e.,  $\partial_\nu f = 0$ ) then becomes

$$\chi = \begin{cases} \frac{T}{\Gamma} \tanh \frac{\Gamma}{T}, & \nu \ll \Gamma^2, \Gamma T \\ 1 - \frac{\Gamma^2 T^2}{\nu^2} - \frac{2\Gamma^2 T^4}{\nu^3}, & \nu \gg T^2, \Gamma T \end{cases} \tag{16}$$

Together with Eq. (11), we find two solutions. The quantum paramagnetic solution is  $\chi \sim \frac{T}{\Gamma} \tanh \frac{\Gamma}{T}$ , with free energy

$$f_Q \sim -T \ln \left(2 \cosh \frac{\Gamma}{T}\right). \tag{17}$$

The classical paramagnetic solution is  $\chi \sim 1 - \frac{4\Gamma^2 T^2}{p^2}$ , with free energy

$$f_C \sim -T \ln 2 - \frac{1}{4T} - \frac{\Gamma^2 T}{p} - \frac{2\Gamma^2 T^3}{p^2}. \tag{18}$$

We determine which is the equilibrium solution by setting  $f_Q = f_C$ . The resulting phase boundary  $\Gamma_b(T)$  is always  $O(1)$ . For  $\Gamma < \Gamma_b(T)$ , the classical paramagnetic free energy is lower. For  $\Gamma > \Gamma_b(T)$ , the quantum paramagnetic free energy is lower.

This calculation is only correct when the temperature is low enough for the equilibrium solution to match the asymptotics we assumed. The classical solution requires  $\nu_C = \frac{1}{2}p \left(1 - \frac{4\Gamma^2 T^2}{p^2}\right)^{p-1} \gg T^2, \Gamma T$ . Since it is the equilibrium solution when  $\Gamma \lesssim O(1)$ , this amounts to  $T \ll \sqrt{p}$ . Similarly, the quantum solution requires  $\nu_Q = \frac{1}{2}p \left(\frac{T}{\Gamma} \tanh \frac{\Gamma}{T}\right)^{p-1} \ll \Gamma^2, \Gamma T$ , which for  $\Gamma \sim O(1)$  amounts to  $T \ll \sqrt{\frac{p}{\ln p}}$ .

For the current paper, we're interested in small  $\Gamma$ , i.e., the classical paramagnetic phase. So we have

$$f = -T \ln 2 - \frac{1}{4T} - \frac{\Gamma^2 T}{p} - \frac{2\Gamma^2 T^3}{p^2} + \dots, \quad (T \ll \sqrt{p}). \tag{19}$$

The energy density at temperature  $T$  is

$$\epsilon(T) = -\frac{1}{2T} + \frac{4\Gamma^2 T^3}{p^2} + \dots, \quad (T \ll \sqrt{p}). \tag{20}$$

The thermodynamic entropy density is

$$s_t(T) = \ln 2 - \frac{1}{4T^2} + \frac{\Gamma^2}{p} + \frac{6\Gamma^2 T^2}{p^2} + \dots, \quad (T \ll \sqrt{p}). \tag{21}$$

Eq. (20) is the relationship between energy and temperature used throughout the main text. Note that we only make use of Eq. (20) for  $T \lesssim \sqrt{\frac{p}{\ln p}}$ , and for those temperatures  $\epsilon(T) \sim -\frac{1}{2T}$  to leading order in  $\frac{1}{p}$ .

### B. Distribution of classical energies.

The classical energy of a configuration  $\alpha$  in the  $p$ -spin model is

$$E_\alpha = \sum_{(i_1 \dots i_p)} J_{i_1 \dots i_p} \sigma_{i_1}^{z(\alpha)} \dots \sigma_{i_p}^{z(\alpha)}, \quad (22)$$

where the sum is over all distinct  $p$ -tuples of  $N$  spins. The couplings  $J_{i_1 \dots i_p}$  are independent Gaussians of mean 0 and variance  $\frac{p!}{2N^{p-1}}$ . Since  $\sigma_{i_1}^{z(\alpha)} \dots \sigma_{i_p}^{z(\alpha)} = \pm 1$  for each  $(i_1 \dots i_p)$ ,  $E_\alpha$  is a sum of Gaussians. Thus it is Gaussian itself, of mean 0 and variance  $\binom{N}{p} \frac{p!}{2N^{p-1}}$ . In terms of the energy density,

$$P_1(\epsilon_\alpha) = A_1 \exp(-N\epsilon_\alpha^2), \quad (23)$$

where  $A_1$  is the normalization.

Now consider the joint distribution for configurations  $\alpha$  and  $\beta$  that differ in the value of  $Nx_{\alpha\beta}$  spins.  $E_\alpha$  and  $E_\beta$  are determined by sums over the same couplings. A given term  $J_{i_1 \dots i_p} \sigma_{i_1}^z \dots \sigma_{i_p}^z$  will contribute the same value for both if an even number of the  $Nx_{\alpha\beta}$  differing spins are in the tuple  $(i_1 \dots i_p)$ . It will contribute opposite values if an odd number are involved. Let  $d_{\text{even}}$  denote the number of  $p$ -tuples that contain an even number of the differing spins, and  $d_{\text{odd}}$  denote the number that contain an odd number. Then  $\frac{E_\alpha + E_\beta}{2}$  is a sum of  $d_{\text{even}}$  Gaussians and  $\frac{E_\alpha - E_\beta}{2}$  is a sum of  $d_{\text{odd}}$  different Gaussians. Note that

$$\begin{aligned} d_{\text{even}} + d_{\text{odd}} &= \sum_{l=0:l \text{ even}}^p \binom{Nx_{\alpha\beta}}{l} \binom{N(1-x_{\alpha\beta})}{p-l} + \sum_{l=1:l \text{ odd}}^p \binom{Nx_{\alpha\beta}}{l} \binom{N(1-x_{\alpha\beta})}{p-l} \\ &= \sum_{l=0}^p \binom{Nx_{\alpha\beta}}{l} \binom{N(1-x_{\alpha\beta})}{p-l} = \binom{N}{p} \sim \frac{N^p}{p!}, \end{aligned} \quad (24)$$

$$\begin{aligned} d_{\text{even}} - d_{\text{odd}} &= \sum_{l=0:l \text{ even}}^p \binom{Nx_{\alpha\beta}}{l} \binom{N(1-x_{\alpha\beta})}{p-l} - \sum_{l=1:l \text{ odd}}^p \binom{Nx_{\alpha\beta}}{l} \binom{N(1-x_{\alpha\beta})}{p-l} \\ &= \sum_{l=0}^p (-1)^l \binom{Nx_{\alpha\beta}}{l} \binom{N(1-x_{\alpha\beta})}{p-l} \\ &\sim \sum_{l=0}^p (-1)^l \frac{N^p}{l!(p-l)!} x_{\alpha\beta}^l (1-x_{\alpha\beta})^{p-l} = \frac{N^p}{p!} (1-2x_{\alpha\beta})^p. \end{aligned} \quad (25)$$

Thus

$$d_{\text{even}} = \frac{N^p}{2p!} (1 + (1-2x_{\alpha\beta})^p), \quad (26)$$

$$d_{\text{odd}} = \frac{N^p}{2p!} (1 - (1-2x_{\alpha\beta})^p). \quad (27)$$

The variance of  $\frac{E_\alpha + E_\beta}{2}$  is  $\frac{N}{4} (1 + (1-2x_{\alpha\beta})^p)$ , and the variance of  $\frac{E_\alpha - E_\beta}{2}$  is  $\frac{N}{4} (1 - (1-2x_{\alpha\beta})^p)$ . Thus the joint distribution of  $\epsilon_\alpha$  and  $\epsilon_\beta$  is

$$P_2(\epsilon_\alpha, \epsilon_\beta) = A_2 \exp \left( -\frac{N}{2} \left( \frac{(\epsilon_\alpha + \epsilon_\beta)^2}{1 + (1-2x_{\alpha\beta})^p} + \frac{(\epsilon_\alpha - \epsilon_\beta)^2}{1 - (1-2x_{\alpha\beta})^p} \right) \right), \quad (28)$$

with  $A_2$  the normalization.

### C. The annealed distance-resolved entropy. Transition temperatures and cluster properties.

Here we consider the following question: if a given configuration  $\alpha$  has energy density  $\epsilon_0$ , how many of the other configurations with energy density  $\epsilon_0$  are at Hamming distance  $Nx$  from  $\alpha$ ? As long as  $\epsilon_0$  is above the ground-state energy density, there will be some configurations having  $\epsilon_0$  with probability 1. Since the configurations are



all statistically equivalent, the fraction of disorder realizations in which  $\alpha$  is localized given it has  $\epsilon_0$  is equal to the fraction of all eigenstates at  $\epsilon_0$  (over all disorder realizations) that are localized. Thus we pick some arbitrary reference configuration and condition on it having energy density  $\epsilon_0$ .

The conditional distribution for a configuration at distance  $x$  is

$$P_x(\epsilon) = \frac{P_2(\epsilon, \epsilon_0)}{P_1(\epsilon_0)} = \frac{A_2}{A_1} e^{-N \frac{(\epsilon - (1-2x)^p \epsilon_0)^2}{1 - (1-2x)^{2p}}}. \quad (29)$$

The expected number of configurations at  $x$  with  $\epsilon_0$  is then  $e^{Ns(x)}$  with

$$\begin{aligned} s(x) &\equiv \frac{1}{N} \ln \mathbb{E} [\#(\alpha \text{ at } x \text{ with } \epsilon_\alpha = \epsilon_0)] \\ &= \frac{1}{N} \ln \binom{N}{Nx} P_x(\epsilon_0) \\ &= -x \ln x - (1-x) \ln(1-x) - \frac{1 - (1-2x)^p}{1 + (1-2x)^p} \epsilon_0^2. \end{aligned} \quad (30)$$

As  $x \rightarrow 0^+$ ,  $s(x) \sim -x \ln x > 0$ . Also,  $s(\frac{1}{2}) = \ln 2 - \epsilon_0^2 = s_t(T)$ , the (annealed) thermodynamic entropy. As long as  $\epsilon_0$  lies within the spectrum,  $s_t(T) > 0$ . Thus  $s(x)$  is positive as  $x$  approaches both endpoints of  $[0, \frac{1}{2}]$ , and there is a transition temperature  $T^*$  at which  $s(x)$  first becomes negative at an intermediate  $x$ .

We can show that  $T^*$  occurs at a scale  $\sqrt{\frac{p}{\ln p}}$  and  $s(x)$  first becomes negative on a scale  $\frac{1}{p\sqrt{\ln p}}$ : set  $x = x_r \frac{1}{p\sqrt{\ln p}}$  and  $\epsilon = \epsilon_r \sqrt{\frac{\ln p}{p}}$ , with  $x_r \sim \Theta(1)$  and  $\epsilon_r \sim \Theta(1)$ . Then

$$\begin{aligned} s(x) &= -\frac{x_r}{p\sqrt{\ln p}} \ln \frac{x_r}{p\sqrt{\ln p}} - \left(1 - \frac{x_r}{p\sqrt{\ln p}}\right) \ln \left(1 - \frac{x_r}{p\sqrt{\ln p}}\right) - \frac{1 - \left(1 - \frac{2x_r}{p\sqrt{\ln p}}\right)^p}{1 + \left(1 - \frac{2x_r}{p\sqrt{\ln p}}\right)^p} \frac{\ln p}{p} \epsilon_r^2 \\ &= -\frac{x_r}{p\sqrt{\ln p}} \ln \frac{x_r}{p\sqrt{\ln p}} + \frac{x_r}{p\sqrt{\ln p}} - \frac{\ln p}{p} \left( \frac{x_r}{\sqrt{\ln p}} - \frac{x_r^3}{3\sqrt{\ln p}^3} \right) \epsilon_r^2 + O\left(\frac{1}{p\sqrt{\ln p}^3}\right) \\ &= \frac{\sqrt{\ln p}}{p} \left( \left(1 + \frac{\ln \ln p}{2 \ln p} - \epsilon_r^2\right) x_r + O\left(\frac{1}{\ln p}\right) \right). \end{aligned} \quad (31)$$

If  $\epsilon_r^2 < 1 + \frac{\ln \ln p}{2 \ln p}$ , then  $s(x)$  increases linearly from 0. Yet if  $\epsilon_r^2 > 1 + \frac{\ln \ln p}{2 \ln p}$ , then  $s(x)$  decreases linearly from 0. This is the transition, which in terms of temperature using Eq. (20) is

$$T^* = \sqrt{\frac{p}{4 \ln p}} \left( 1 + O\left(\frac{\ln \ln p}{\ln p}\right) \right). \quad (32)$$

Below  $T^*$ , there is an initial region (“cluster”) of configurations at small  $x$  and a separate region of configurations at larger  $x$ . These are separated by a region  $(x^*(T), x^{**}(T))$  in which none of the configurations have energy density  $\epsilon_0(T)$ . Assuming  $x \ll \frac{1}{p}$ ,

$$s(x) \sim -x \ln x + x - p \epsilon_0^2 x. \quad (33)$$

We see that  $s(x) > 0$  for  $x < x^*(T)$ , with

$$x^*(T) = e^{1 - p \epsilon_0^2}. \quad (34)$$

This is the length of a typical cluster. As a consistency check, note that  $x^*(T)$  is indeed much smaller than  $\frac{1}{p}$  so long as  $T \ll T^*$ .

For  $x \gg \frac{1}{p}$ ,

$$s(x) \sim -x \ln x - (1-x) \ln(1-x) - \epsilon_0^2, \quad (35)$$

to within exponentially small corrections.  $x^{**}(T)$  is the root of this equation, which is  $O(1)$  when  $\epsilon_0^2 \sim O(1)$ . Yet for small  $\epsilon_0$ ,

$$x^{**}(T) \sim \frac{\epsilon_0^2}{\ln \frac{1}{\epsilon_0^2}}. \quad (36)$$

Note that  $x^{**}(T) \gg \frac{1}{p}$  so long as  $T \ll T^*$ .

Thus  $x^{**}(T) - x^*(T) \gg x^*(T)$  when  $T \ll T^*$ . In words, the separation between clusters is much larger than the length of a cluster. This is not true above  $T^*$ , where one cannot distinguish different clusters.

Since the expected number of configurations at  $x$  is  $e^{Ns(x)}$ , the total number of configurations within a cluster is dominated by the maximum of  $s(x)$ , which is

$$\max_{x < x^*(T)} s(x) = e^{-p\epsilon_0^2}. \quad (37)$$

#### D. Replica calculation of the forward-scattering wavefunction.

From the main text, the forward-scattering expression for an eigenstate  $|\Psi\rangle$  of the quantum  $p$ -spin model is

$$\langle \beta | \Psi \rangle = \frac{\Gamma}{E_\alpha - E_\beta} \sum_{\mathcal{P}} \prod_{\gamma_j \in \mathcal{P}} \frac{\Gamma}{E_\alpha - E_{\gamma_j}}. \quad (38)$$

Here  $\alpha$  and  $\beta$  are classical configurations of energy  $E_\alpha$  and  $E_\beta$ , separated by Hamming distance  $Nx_{\alpha\beta}$ . At  $\Gamma = 0$ ,  $|\Psi\rangle = |\alpha\rangle$ . The sum is over all  $(Nx_{\alpha\beta})!$  direct “paths” (i.e., sequences of spin flips) from configuration  $\alpha$  to configuration  $\beta$ , and  $\gamma_j$  is the  $j$ ’th configuration along path  $\mathcal{P}$ . From now on we’ll take  $E_\alpha \equiv E_0$  and  $x_{\alpha\beta} \equiv x$ . We’ll also assume that the statistics of  $\langle \beta | \Psi \rangle$ , having very heavy tails, aren’t significantly affected by cancellations among paths. Thus we make the analytically expedient replacement

$$|\langle \beta | \Psi \rangle| \rightarrow \frac{\Gamma}{|E_0 - E_\beta|} \sum_{\mathcal{P}} \prod_{\gamma_j \in \mathcal{P}} \frac{\Gamma}{|E_0 - E_{\gamma_j}|}. \quad (39)$$

In the main text, we assumed that all paths leading up to  $\beta$  have typical amplitudes, and then determined the probability that  $|E_0 - E_\beta|$  is small enough to make  $|\langle \beta | \Psi \rangle| > 1$ . Here “typical” means replacing each  $E_{\gamma_j}$  by its mean value, which from Eq. (29) is  $(1 - \frac{2j}{N})^p E_0$  (note that configuration  $\gamma_j$  by definition differs from  $\alpha$  by  $j$  spin-flips). We find that

$$|\langle \beta | \Psi \rangle|_{\text{typ}} = \frac{\Gamma}{|E_0 - E_\beta|} (Nx)! \prod_{j=1}^{Nx-1} \frac{\Gamma}{|E_0| \left(1 - \left(1 - \frac{2j}{N}\right)^p\right)} \equiv \frac{1}{|E_0 - E_\beta|} e^{-Nk(x)}, \quad (40)$$

with

$$k(x) \sim -x \ln \frac{x\Gamma}{e|\epsilon_0|} + \int_0^x dy \ln(1 - (1 - 2y)^p). \quad (41)$$

Note that in the main text, we had instead used the notation  $|\langle \beta | \Psi \rangle|_{\text{typ}} = \frac{1}{|E_0 - E_\beta|} (Nx)! e^{-Nk(x)}$ .

Here we go beyond the typical amplitudes by making a replica-symmetry-breaking ansatz, to see whether the typical value of the sum in Eq. (39) is approximated by the sum of the typical values. We compute  $\frac{1}{N} \mathbb{E} \ln |\langle \beta | \Psi \rangle| = \frac{1}{Nn} \lim_{n \rightarrow 0} \ln \mathbb{E} |\langle \beta | \Psi \rangle|^n$  assuming one step of replica-symmetry-breaking. The  $n$  replicated paths cluster into  $\frac{n}{m}$  independent groups of  $m$  identical paths each:

$$\begin{aligned} \mathbb{E} |\langle \beta | \Psi \rangle|^n &= \frac{\Gamma^{Nnx}}{N^{Nnx}} \sum_{\mathcal{P}_1 \dots \mathcal{P}_n} \mathbb{E} \left( \prod_{\gamma_j \in \mathcal{P}_1} \frac{1}{|\epsilon_0 - \epsilon_{\gamma_j}|} \right) \dots \left( \prod_{\gamma_j \in \mathcal{P}_n} \frac{1}{|\epsilon_0 - \epsilon_{\gamma_j}|} \right) \\ &\rightarrow \frac{\Gamma^{Nnx}}{N^{Nnx}} \sum_{\mathcal{P}_1 \dots \mathcal{P}_{\frac{n}{m}}} \mathbb{E} \left( \prod_{\gamma_j \in \mathcal{P}_1} \frac{1}{|\epsilon_0 - \epsilon_{\gamma_j}|^m} \right) \dots \mathbb{E} \left( \prod_{\gamma_j \in \mathcal{P}_{\frac{n}{m}}} \frac{1}{|\epsilon_0 - \epsilon_{\gamma_j}|^m} \right) \end{aligned} \quad (42)$$

The second line is our replica-symmetry-breaking ansatz. Recall that within this framework,  $0 \leftarrow n < m \leq 1$ . To make further progress, we replace

$$\mathbb{E} \left( \prod_{\gamma_j \in \mathcal{P}} \frac{1}{|\epsilon_0 - \epsilon_{\gamma_j}|^m} \right) \rightarrow \prod_{\gamma_j \in \mathcal{P}} \mathbb{E} \frac{1}{|\epsilon_0 - \epsilon_{\gamma_j}|^m}, \quad (43)$$

even though the energies along a path are correlated. Although this isn't a controlled approximation, it does still let us probe how rare fluctuations might dominate the sum. Denoting  $x_j \equiv \frac{j}{N}$ ,

$$\begin{aligned}
\mathbb{E} \frac{1}{|\epsilon_0 - \epsilon_{\gamma_j}|^m} &= \int_{-\infty}^{\infty} d\epsilon \sqrt{\frac{N}{\pi(1 - (1 - 2x_j)^{2p})}} e^{-N \frac{(\epsilon - (1 - 2x_j)^p \epsilon_0)^2}{1 - (1 - 2x_j)^{2p}}} \frac{1}{|\epsilon_0 - \epsilon|^m} \\
&= \sqrt{\frac{N\epsilon_0^2}{\pi(1 - (1 - 2x_j)^{2p})}} |\epsilon_0|^{-m} \int_{-\infty}^{\infty} dz \frac{1}{|1 - z|^m} e^{-N\epsilon_0^2 \frac{(z - (1 - 2x_j)^p)^2}{1 - (1 - 2x_j)^{2p}}} \\
&= \sqrt{\frac{N\epsilon_0^2}{\pi(1 - (1 - 2x_j)^{2p})}} |\epsilon_0|^{-m} \left( \int_{-\infty}^1 dz \frac{1}{(1 - z)^m} e^{-N\epsilon_0^2 \frac{(z - (1 - 2x_j)^p)^2}{1 - (1 - 2x_j)^{2p}}} \right. \\
&\quad \left. + \int_1^{\infty} dz \frac{1}{(z - 1)^m} e^{-N\epsilon_0^2 \frac{(z - (1 - 2x_j)^p)^2}{1 - (1 - 2x_j)^{2p}}} \right).
\end{aligned} \tag{44}$$

Two things are happening with this expression. The first integral has a saddle-point contribution:

$$\mathbb{E} \frac{1}{|\epsilon_0 - \epsilon_{\gamma_j}|^m} \rightarrow \frac{1}{(1 - (1 - 2x_j)^p)^m |\epsilon_0|^m} + O\left(\frac{1}{N}\right). \tag{45}$$

However, this expression does not diverge as  $m \rightarrow 1^-$ , yet we know the full expression does. We must be careful that the diverging part of the integral doesn't get buried in a remainder. For that, we integrate by parts:

$$\begin{aligned}
\int_{-\infty}^1 dz \frac{1}{(1 - z)^m} e^{-N\epsilon_0^2 \frac{(z - (1 - 2x_j)^p)^2}{1 - (1 - 2x_j)^{2p}}} &= \frac{2N\epsilon_0^2}{1 - (1 - 2x_j)^{2p}} \frac{1}{1 - m} \int_{-\infty}^1 dz (1 - z)^{1-m} ((1 - 2x_j)^p - z) e^{-N\epsilon_0^2 \frac{(z - (1 - 2x_j)^p)^2}{1 - (1 - 2x_j)^{2p}}} \\
&\rightarrow \frac{1}{1 - m} e^{-N\epsilon_0^2 \frac{(1 - (1 - 2x_j)^p)^2}{1 - (1 - 2x_j)^{2p}}} + O((1 - m)^0),
\end{aligned} \tag{46}$$

and similarly for the second integral. Putting Eqs. (45) and (46) together, we have that

$$\mathbb{E} \frac{1}{|\epsilon_0 - \epsilon_{\gamma_j}|^m} = \frac{1}{(1 - (1 - 2x_j)^p)^m |\epsilon_0|^m} \left( 1 + \sqrt{\frac{N\epsilon_0^2 b_j}{\pi}} \frac{2}{1 - m} e^{-N\epsilon_0^2 b_j} \right) + O\left(\frac{(1 - m)^0}{N}\right), \tag{47}$$

where we've defined  $b_j \equiv \frac{1 - (1 - 2x_j)^p}{1 + (1 - 2x_j)^p}$ .

Eq. (47) goes into Eq. (42):

$$\begin{aligned}
\mathbb{E} |\langle \beta | \Psi \rangle|^n &\sim \frac{\Gamma^{Nnx}}{N^{Nnx}} (Nx)!^{\frac{n}{m}} \prod_{j=1}^{Nx} \frac{1}{(1 - (1 - 2x_j)^p)^n |\epsilon_0|^n} \left( 1 + \sqrt{\frac{N\epsilon_0^2 b_j}{\pi}} \frac{2}{1 - m} e^{-N\epsilon_0^2 b_j} \right)^{\frac{n}{m}} \\
&\sim \frac{\Gamma^{Nnx}}{(N|\epsilon_0|)^{Nnx}} e^{-n \sum_{j=1}^{Nx} \ln(1 - (1 - 2x_j)^p)} \left( \frac{Nx}{e} \right)^{\frac{Nnx}{m}} e^{\frac{n}{m} \sum_{j=1}^{Nx} \ln \left( 1 + \sqrt{\frac{N\epsilon_0^2 b_j}{\pi}} \frac{2}{1 - m} e^{-N\epsilon_0^2 b_j} \right)} \\
&\sim \exp \left( Nn \left( \phi_0(x) + \frac{1}{m} \phi_1(x, m) \right) \right),
\end{aligned} \tag{48}$$

with

$$\phi_0(x) = x \ln \frac{\Gamma}{N|\epsilon_0|} - \int_0^x dy \ln(1 - (1 - 2y)^p), \tag{49}$$

$$\phi_1(x, m) = x \ln \frac{Nx}{e} + \frac{1}{N} \sum_{j=1}^{Nx} \ln \left( 1 + \sqrt{\frac{N\epsilon_0^2 b_j}{\pi}} \frac{2}{1 - m} e^{-N\epsilon_0^2 b_j} \right). \tag{50}$$

We now *minimize* with respect to  $m$  (recall that within the replica formalism we minimize instead of maximize). The minimum exists at a non-trivial value  $m_{\text{EQ}}$  because  $\frac{1}{m} \phi_1(x, m) \rightarrow +\infty$  as  $m \rightarrow 0$  and as  $m \rightarrow 1$ . We need to

solve  $\frac{1}{m}\partial_m\phi_1 - \frac{1}{m^2}\phi_1 = 0$ , which can be written

$$Nx \ln \frac{Nx}{e} + \sum_{j=1}^{Nx} \ln \left( 1 + \sqrt{\frac{N\epsilon_0^2 b_j}{\pi}} \frac{2}{1 - m_{\text{EQ}}} e^{-N\epsilon_0^2 b_j} \right) = \frac{m_{\text{EQ}}}{1 - m_{\text{EQ}}} \sum_{j=1}^{Nx} \frac{\sqrt{\frac{N\epsilon_0^2 b_j}{\pi}} \frac{2}{1 - m_{\text{EQ}}} e^{-N\epsilon_0^2 b_j}}{1 + \sqrt{\frac{N\epsilon_0^2 b_j}{\pi}} \frac{2}{1 - m_{\text{EQ}}} e^{-N\epsilon_0^2 b_j}}. \quad (51)$$

It would be very difficult to fully solve this equation for  $m_{\text{EQ}}$ , but we can pull out how  $1 - m_{\text{EQ}} \equiv \Delta$  scales with  $N$ . We'll find that  $\Delta \ll 1$ . First consider

$$B_j \equiv \sqrt{\frac{N\epsilon_0^2 b_j}{\pi}} \frac{2}{\Delta} e^{-N\epsilon_0^2 b_j}. \quad (52)$$

$B_j \gg 1$  (as  $\Delta \rightarrow 0$ ) when  $b_j \lesssim \frac{1}{N\epsilon_0^2} \ln \frac{1}{\Delta}$ .  $B_j$  is small otherwise. Assuming  $\ln \frac{1}{\Delta} \ll N$  (we'll see that this turns out to be true),  $b_j \lesssim \frac{1}{N\epsilon_0^2} \ln \frac{1}{\Delta}$  corresponds to  $j \lesssim \frac{1}{p\epsilon_0^2} \ln \frac{1}{\Delta}$ . The maximum of  $B_j$  is  $\sqrt{\frac{2}{\pi e}} \frac{1}{\Delta}$ , at  $j = \frac{1}{2p\epsilon_0^2}$ . Thus we have the bounds

$$\frac{C_1}{p\epsilon_0^2} \ln \frac{1}{\Delta} < \sum_{j=1}^{Nx} \ln \left( 1 + \sqrt{\frac{N\epsilon_0^2 b_j}{\pi}} \frac{2}{\Delta} e^{-N\epsilon_0^2 b_j} \right) < \frac{C_2}{p\epsilon_0^2} \left( \ln \frac{1}{\Delta} \right)^2, \quad (53)$$

for suitable choices of the constants  $C_1$  and  $C_2$ . Since  $\ln \frac{1}{\Delta} \ll N$ , we see that this sum is negligible compared to  $Nx \ln \frac{Nx}{e}$  in Eq. (51). Similarly, the RHS of Eq. (51) is bounded by

$$\frac{D_1}{p\epsilon_0^2} \frac{1}{\Delta} \ln \frac{1}{\Delta} < \frac{1 - \Delta}{\Delta} \sum_{j=1}^{Nx} \frac{\sqrt{\frac{N\epsilon_0^2 b_j}{\pi}} \frac{2}{\Delta} e^{-N\epsilon_0^2 b_j}}{1 + \sqrt{\frac{N\epsilon_0^2 b_j}{\pi}} \frac{2}{\Delta} e^{-N\epsilon_0^2 b_j}} < \frac{D_2}{p\epsilon_0^2} \frac{1}{\Delta} \ln \frac{1}{\Delta}, \quad (54)$$

for suitable choices of the constants  $D_1$  and  $D_2$ . Since  $\Delta$  is chosen so that this sum is essentially  $Nx \ln \frac{Nx}{e}$ ,

$$\frac{D_1}{p\epsilon_0^2} \frac{1}{\Delta} \ln \frac{1}{\Delta} < Nx \ln \frac{Nx}{e} < \frac{D_2}{p\epsilon_0^2} \frac{1}{\Delta} \ln \frac{1}{\Delta}. \quad (55)$$

We see that  $\Delta \sim \frac{1}{Nx}$ , so that

$$m_{\text{EQ}} = 1 - \frac{\delta(Nx, p\epsilon_0^2)}{Nx}, \quad (56)$$

with  $\delta(\cdot, \cdot)$  an  $O(1)$  function. Although  $\delta(Nx, p\epsilon_0^2)$  does vary with  $N$ , it stays  $O(1)$  as  $N \rightarrow \infty$ .

Thus

$$\frac{1}{m_{\text{EQ}}} \phi_1(x, m_{\text{EQ}}) = x \ln \frac{Nx}{e} + O\left(\frac{(\ln N)^2}{N}\right). \quad (57)$$

Our final result for the average log-amplitude, which is to be compared with the “typical” estimate in Eqs. (40) and (41), is

$$\begin{aligned} \frac{1}{N} \mathbb{E} \ln |\langle \beta | \Psi \rangle| &= \frac{1}{Nn} \lim_{n \rightarrow 0} \ln \mathbb{E} |\langle \beta | \Psi \rangle|^n = \phi_0(x) + \frac{1}{m_{\text{EQ}}} \phi_1(x, m_{\text{EQ}}) \\ &= x \ln \frac{x\Gamma}{e|\epsilon_0|} - \int_0^x dy \ln(1 - (1 - 2y)^p) + O\left(\frac{(\ln N)^2}{N}\right). \end{aligned} \quad (58)$$

We see that the corrections to Eq. (41) vanish in the thermodynamic limit, even when allowing for one step of replica-symmetry-breaking. This justifies our use of the “typical” estimate, denoted  $k(x)$ , in the main text.

### E. MBL phase boundary.

As justified in the previous section, the amplitude of eigenstate  $|\Psi\rangle$  on configuration  $|\beta\rangle$  is

$$|\langle \beta | \Psi \rangle| = \frac{1}{|E_0 - E_\beta|} e^{-Nk(x)}, \quad (59)$$

with  $k(x)$  given by Eq. (41). Here  $E_0$  is the energy of the eigenstate, related to temperature by Eq. (20), and  $Nx$  is the Hamming distance from  $\beta$  to the unperturbed ( $\Gamma = 0$ ) configuration. Assuming  $e^{-Nk(x)} \ll 1$  (which it is for all  $T$  and  $\Gamma$  of interest),  $|E_0 - E_\beta|$  must be exponentially small for  $\beta$  to be resonant. The distribution of  $E_\beta$  is essentially uniform on such small scales, so a fraction  $e^{-Nk(x)}$  of the configurations at distance  $x$  with energy density  $\epsilon_0$  resonate. There are  $e^{Ns(x)}$  such configurations, with  $s(x)$  given by Eq. (30), so the average number of resonant configurations at distance  $x$  is  $e^{Nf(x)}$ , with

$$f(x) = s(x) - k(x) = x \ln \frac{\Gamma}{e|\epsilon_0|} - (1-x) \ln(1-x) - \int_0^x dy \ln(1 - (1-2y)^p) - \frac{1 - (1-2x)^p}{1 + (1-2x)^p} \epsilon_0^2. \quad (60)$$

If  $f(x) < 0$ , then the probability of having *any* resonance at distance  $x$  is 0 in the thermodynamic limit. If  $f(x) > 0$ , then there are resonances at distance  $x$  and the eigenstate  $|\Psi\rangle$  is hybridized among them. The forward-scattering expression Eq. (38) is then invalid, yet one could in principle write down an analogous expression in terms of “paths” among hybridized states. These states would then hybridize if further resonances are found, etc.

For now, consider  $p$  large and  $x$  independent of  $p$ . Then  $(1-2x)^p \rightarrow 0$  and

$$\begin{aligned} \int_0^x dy \ln(1 - (1-2y)^p) &= \frac{1}{p} \int_0^{px} dz \ln\left(1 - \left(1 - \frac{2z}{p}\right)^p\right) \\ &= \frac{1}{p} \int_0^\infty dz \ln(1 - e^{-2z}) + O\left(\frac{1}{p^2}\right) = -\frac{\pi^2}{12p} + O\left(\frac{1}{p^2}\right). \end{aligned} \quad (61)$$

Thus

$$f(x) = x \ln \frac{\Gamma}{e|\epsilon_0|} - (1-x) \ln(1-x) - \epsilon_0^2 + \frac{\pi^2}{12p} + O\left(\frac{1}{p^2}\right). \quad (62)$$

Note that this expression is valid only for  $x \sim O(1)$  with respect to  $\frac{1}{p}$ . We recover the expression previously obtained for the REM as  $p \rightarrow \infty$  [24]. As done in Ref. [24], we identify any eigenstate with  $f(x) > 0$  for some  $x \sim O(1)$  as delocalized. We consider the case of  $x \ll 1$  separately. The maximum of Eq. (62) is at  $x = 1 - \frac{|\epsilon_0|}{\Gamma}$ , with a value of

$$f_{\max} = \ln \frac{\Gamma}{|\epsilon_0|} - 1 + \frac{|\epsilon_0|}{\Gamma} - \epsilon_0^2 + \frac{\pi^2}{12p} + O\left(\frac{1}{p^2}\right). \quad (63)$$

$\Gamma_c(\epsilon_0)$  (expressed as  $\Gamma_c(T)$  in the main text) is defined by  $f_{\max} = 0$ . For small  $\epsilon_0$ ,

$$\Gamma_c(\epsilon_0) = |\epsilon_0| \left(1 + \sqrt{2\epsilon_0^2 - \frac{\pi^2}{6p}} + O(\epsilon_0^2)\right) \Leftrightarrow \Gamma_c(T) = \frac{1}{2T} \left(1 + \sqrt{\frac{2}{T^2} - \frac{\pi^2}{6p}} + O\left(\frac{1}{T^2}\right)\right). \quad (64)$$

The corrections to the REM phase boundary are only  $O\left(\frac{1}{p}\right)$ , yet we see that for any fixed  $p$ , this expression breaks down at  $|\epsilon_0| < \frac{\pi}{\sqrt{12p}}$ . In fact, problems emerge outside of this range as well, due to the behavior of  $f(x)$  at small  $x$ .

For  $x \ll \frac{1}{p}$ ,

$$f(x) = x \ln \frac{\Gamma}{2p|\epsilon_0|x} + x - p\epsilon_0^2 x + O(p^2 x^2). \quad (65)$$

This expression is positive for  $x < x_{\text{res}}^*$ , where

$$x_{\text{res}}^* = \frac{\Gamma}{2p|\epsilon_0|} e^{1-p\epsilon_0^2}. \quad (66)$$

Compare to  $x^*$  in Eq. (34). Thus there are always resonances at small enough  $x$ , which lie in the local cluster of configurations at  $x < x^*$ . Suppose for now that  $\epsilon_0 \sim O(1)$ , so that  $x_{\text{res}}^*$  and  $x^*$  are exponentially small in  $p$ . We have no way of describing the wavefunction amplitude on these resonant configurations, and restarting the perturbation series around them will undoubtedly put additional configurations into resonance. But regardless of how the wavefunction hybridizes, there cannot be any resonances at  $x \in (x^*, x^{**})$ , simply because none of those configurations have energy density  $\epsilon_0$ . Thus the entropy of resonances  $f(x)$  at  $x > x^{**}$  is only corrected by an exponentially small amount if we start the forward-scattering at *any* short-distance resonance, even under the worst-case assumption that the resonances extend to  $x^*$ . And furthermore, we show in the section below that the properties of localized eigenstates

are controlled by the amplitude at distances much larger than  $x^*$ , so the short-distance resonances do not affect observables either. All this is to say that, at least for  $\epsilon_0 \sim O(1)$ , none of our conclusions above or in the main text are affected by the short-distance resonances that we're unable to correct for.

However, this reasoning relies on the separation between clusters being much larger than the length of a cluster, which breaks down when  $|\epsilon_0| \lesssim \sqrt{\frac{\ln p}{p}}$ . Thus our description of the localized eigenstates, in particular Eq. (64) and the section below, only applies to those with  $|\epsilon_0| \gg \sqrt{\frac{\ln p}{p}}$ .

### F. Observables in the localized phase.

The localized eigenstates have a highly-correlated “core” that extends no farther than  $x^*$ , and also a tail in which the wavefunction amplitudes are typical. Any amplitude at distance  $x > x^*$  must lie between  $e^{-Nk(x+x^*)}$  and  $e^{-Nk(x-x^*)}$  (see Eq. (40)), since these are the extreme locations within the core from which one could forward-scatter. Assuming  $x \gg x^*$ , the leading behavior of the exponent is simply  $-k(x)$ . Then the total weight in the tail of the wavefunction is

$$\int_{x^*}^1 dx \binom{N}{Nx} \left( e^{-Nk(x)} \right)^2 = \int_{x^*}^1 dx \exp \left( N \left( x \ln \frac{x\Gamma^2}{e^2 \epsilon_0^2} - (1-x) \ln(1-x) - 2 \int_0^x dy \ln(1 - (1-2y)^p) \right) \right). \quad (67)$$

The saddle-point is at  $x_w \sim \frac{\Gamma^2}{4p^2 \epsilon_0^2}$ . For now, assume that  $\Gamma$  is not exponentially small in  $p$ , so that  $x_w$  is indeed much larger than  $x^*$ . The total weight in the tail scales as

$$\exp \left( N \frac{\Gamma^2}{4p^2 \epsilon_0^2} \right). \quad (68)$$

Even if we assume every configuration in the core has amplitude 1 (i.e., as much weight as the unperturbed state), the core cannot contribute more weight than  $\binom{N}{Nx^*} \sim \exp(Np\epsilon_0^2 e^{1-p\epsilon_0^2}) \ll \exp \left( N \frac{\Gamma^2}{4p^2 \epsilon_0^2} \right)$ . Thus the tail of the wavefunction contains all the weight. This allows us to compute observables.

The inverse participation ratio (IPR) of an eigenstate is defined as

$$Y_2(\Psi) = \sum_{\alpha} \frac{|\langle \alpha | \Psi \rangle|^4}{\left( \sum_{\beta} |\langle \beta | \Psi \rangle|^2 \right)^2}. \quad (69)$$

The IPR quantifies how many basis states  $|\alpha\rangle$  the wavefunction  $|\Psi\rangle$  is spread over. In our case, the Hilbert space has dimension  $2^N$ . A wavefunction with  $Y_2 = 2^{-N}$  has equal weight on all basis states, whereas a wavefunction with  $Y_2 = 1$  only has weight on a single basis state. As before, we find that  $\sum_{\alpha} |\langle \alpha | \Psi \rangle|^4$  is dominated by states in the tail:

$$\begin{aligned} \int_{x^*}^1 dx \binom{N}{Nx} \left( e^{-Nk(x)} \right)^4 &= \int_{x^*}^1 dx \exp \left( N \left( x \ln \frac{x^3 \Gamma^4}{e^4 \epsilon_0^4} - (1-x) \ln(1-x) - 4 \int_0^x dy \ln(1 - (1-2y)^p) \right) \right) \\ &\sim \exp \left( N \frac{\Gamma^4}{16p^4 \epsilon_0^4} \right) \end{aligned} \quad (70)$$

Thus

$$Y_2 \sim \exp \left( -N \frac{\Gamma^2}{2p^2 \epsilon_0^2} \left( 1 - \frac{\Gamma^2}{8p^2 \epsilon_0^2} \right) \right). \quad (71)$$

Compare  $\ln Y_2^{-1}$  with the thermodynamic entropy (Eq. (21)). We see that although the localized wavefunction spreads over exponentially many configurations, it still covers an exponentially small fraction of the total number at  $\epsilon_0$ .

We next consider the quantity  $q_{\text{ES}}(\Psi)$  defined as

$$q_{\text{ES}}(\Psi) = \frac{1}{N} \sum_i \left( \frac{\langle \Psi | \hat{\sigma}_i^z | \Psi \rangle}{\langle \Psi | \Psi \rangle} \right)^2. \quad (72)$$

This is an eigenstate analogue to the Edwards-Anderson order parameter  $q_{\text{EA}} \equiv \frac{1}{N} \sum_i \langle \sigma_i^z \rangle^2$ . Whereas the average in  $q_{\text{EA}}$  is over the Gibbs distribution, in Eq. (72) we average over the distribution defined by the eigenstate  $|\Psi\rangle$ .

$q_{\text{ES}}$  gives the overlap between two independent measurements of the  $\sigma^z$  configuration within  $|\Psi\rangle$ . To see this, define  $c_\alpha \equiv \frac{\langle \alpha | \Psi \rangle}{\sqrt{\langle \Psi | \Psi \rangle}}$ , where  $|\alpha\rangle$  is a  $\sigma^z$  basis state. Then

$$q_{\text{ES}}(\Psi) = \frac{1}{N} \sum_i \left( \sum_\alpha \sigma_i^z(\alpha) |c_\alpha|^2 \right)^2 = \sum_{\alpha, \beta} |c_\alpha|^2 |c_\beta|^2 \left( \frac{1}{N} \sum_i \sigma_i^z(\alpha) \sigma_i^z(\beta) \right). \quad (73)$$

The quantity  $\frac{1}{N} \sum_i \sigma_i^z(\alpha) \sigma_i^z(\beta) \equiv Q_{\alpha\beta}$  is the spin overlap between configurations  $\alpha$  and  $\beta$ , exactly as in mean-field spin glass theory.

To evaluate  $q_{\text{ES}}$  in a localized eigenstate, first note that  $Q_{\alpha\beta} = 1 - 2x_{\alpha\beta}$ , with  $x_{\alpha\beta}$  the fractional Hamming distance between  $\alpha$  and  $\beta$ . Then

$$q_{\text{ES}} = e^{-N \frac{\Gamma^2}{2p^2 \epsilon_0^2}} \int dx \binom{N}{Nx} \int dy \int_{|x-y|}^{x+y} dz \binom{Nx}{N \frac{x-y+z}{2}} \binom{N(1-x)}{N \frac{-x+y+z}{2}} e^{-2Nk(x)} e^{-2Nk(y)} (1-2z). \quad (74)$$

Here  $x$  is the distance to  $\alpha$  from the origin,  $y$  is the distance to  $\beta$  from the origin, and  $z$  is the distance from  $\alpha$  to  $\beta$ . Note that  $z$  must be between  $|x-y|$  and  $x+y$ , and the number of  $\beta$  that satisfy this geometry for each  $\alpha$  is  $\binom{Nx}{N \frac{x-y+z}{2}} \binom{N(1-x)}{N \frac{-x+y+z}{2}}$ . The saddle-point in  $z$  is at  $z^* = x+y-2xy$ , and  $\binom{Nx}{N \frac{x-y+z^*}{2}} \binom{N(1-x)}{N \frac{-x+y+z^*}{2}} = \binom{N}{Ny}$  (to leading order). We then have separate saddle-point integrals over  $x$  and  $y$ , both of which are dominated by  $x_w = \frac{\Gamma^2}{4p^2 \epsilon_0^2}$ . The end result is that

$$q_{\text{ES}} = 1 - \frac{\Gamma^2}{p^2 \epsilon_0^2} + \dots \quad (75)$$

As mentioned above, these estimates do implicitly assume that  $\Gamma \gg e^{-p\epsilon_0^2}$ . If  $\Gamma$  is exponentially small in  $p$ , then the bulk of the wavefunction amplitude is in the core and these estimates do not apply. However, the wavefunctions are still localized, and the qualitative results in this section hold:  $\ln Y_2^{-1}$  is much smaller than the thermodynamic entropy, and  $q_{\text{ES}}$  is close to 1.

---